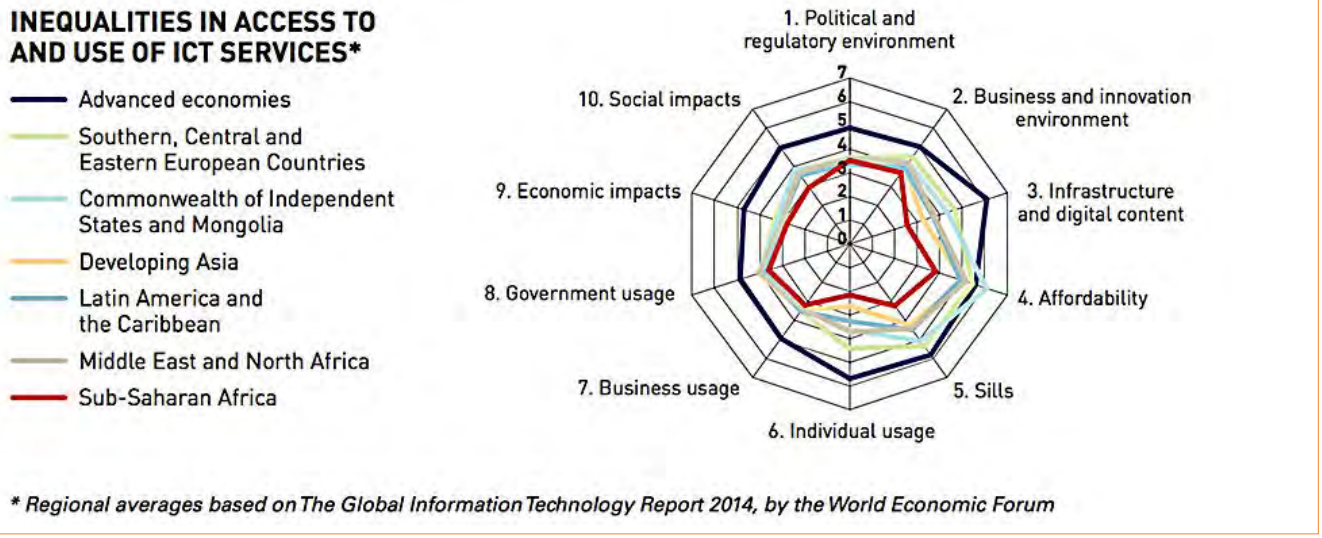
PART 2: BIG DATA – A changing world

Dr Susan Bliss

'Open Data' in a Big Data world

Open Data refers to everyone, everywhere, everyday possessing access to 'all' data. This aims to prevent discrimination and promote transparency and accountability. Of the 4.3 billion people without access to the internet, 90% live in developing countries, mostly in rural remote locations. Access and use of ICT is required to penetrate remote areas in all countries to enable sustainable development.

Spider graph: inequalities in access to and use of ICT services



Source: http://www.undatarevolution.org/wp-content/uploads/2014/10/Screen-Shot-2014-10-24-at-10.38.01-AM.png

However not 'all data' is available to 'all' people because:

- firewalls are built to monitor and control network traffic and block computer hackers
- black holes exist when internet traffic is silently discarded e.g. China and North Korea
- unequal access to data and technologies across regions-advanced/developed economies tend to be ahead of the rest of the world on almost every ICT indicator:
  - access – affordability, infrastructure, content, skills
  - use – government, business, individual
  - Impacts – social, economic, environmental

Governments are opening their data, as Open Data aims to stimulate innovation, tackle economic problems and improve human wellbeing. The Australian Government developed a Big Data strategy to make data held by national and regional authorities publicly available

Firewall concept

http://s3.amazonaws.com/rv-wp/centurylinkdealscominternetres/uploads/2013/11/firewall.jpg

'Better Data Better Lives': 2030 Sustainable Development Goals

Big-Open Data is important for the success of the United Nations(UN) 2030 Sustainable Development Goals (SDG). Huge data sets, supported by technical and analytical services, are required to address SDGs-17 goals, 169 target and 230 indicators.

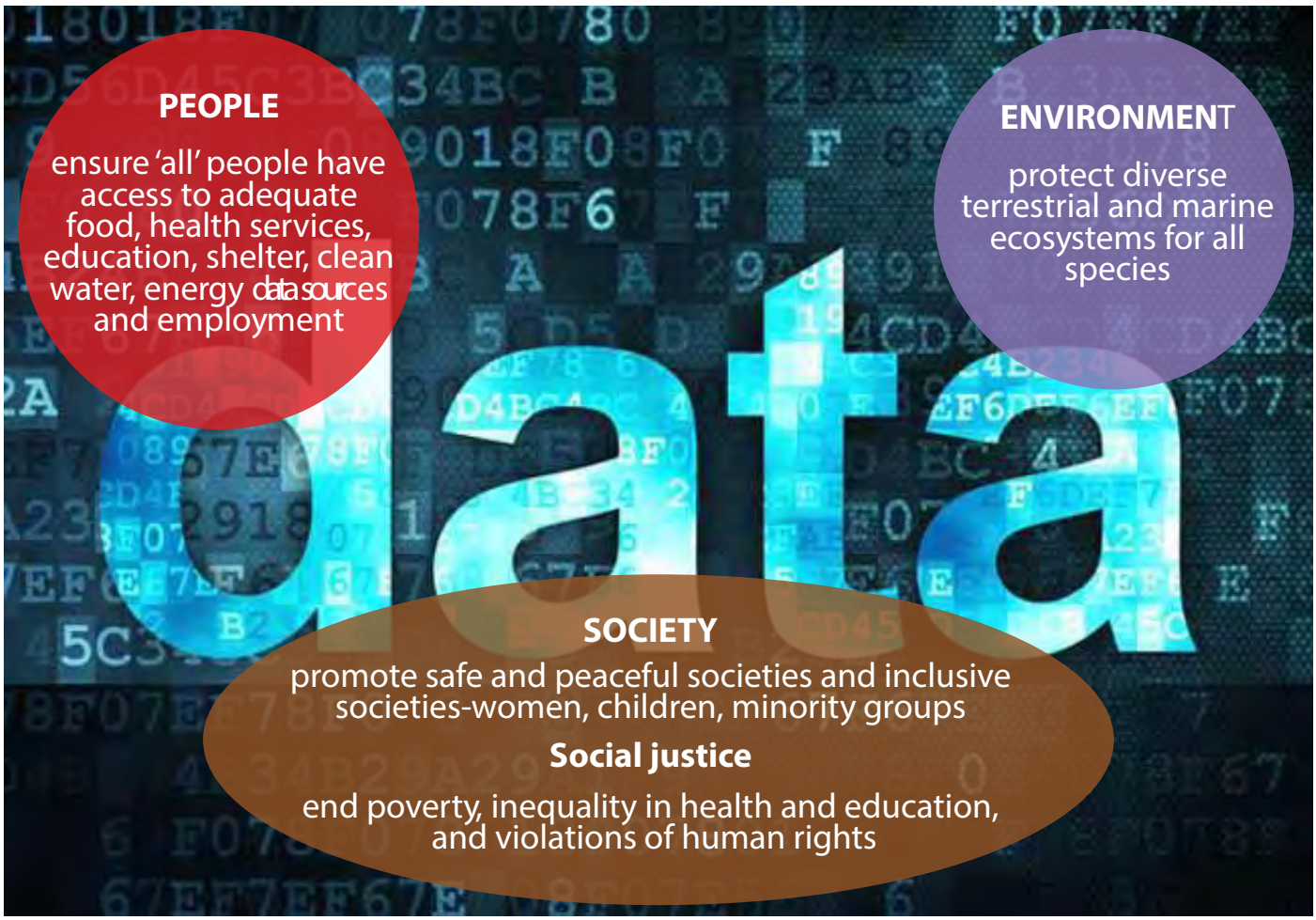
The UN Global Partnership for Sustainable Development Data (GPSDD) is an open, multi-stakeholder network committed to harnessing Big-Open data for sustainable development.

- Open Data: The SDGs pledge to 'leave no one behind', requires data to cover populations, areas and themes, not previously counted or studied.
- Big Data: Countries are required to improve statistical systems and use quality data to improve development policies and monitor development processes.

Big-Open Data underpins the success of UN development policies by focussing on key questions such as:

- Where will scarce funds and resources be allocated? (countries/regions)
- What areas of development require highest assistance? (health/education/water)
- What changes have occurred since the UN intervention? (increase/stagnation/decrease) Aimed to improve human wellbeing (people, societies) and protect the environment, development data will need to be accessible and useable, and to be used more effectively and efficiently.

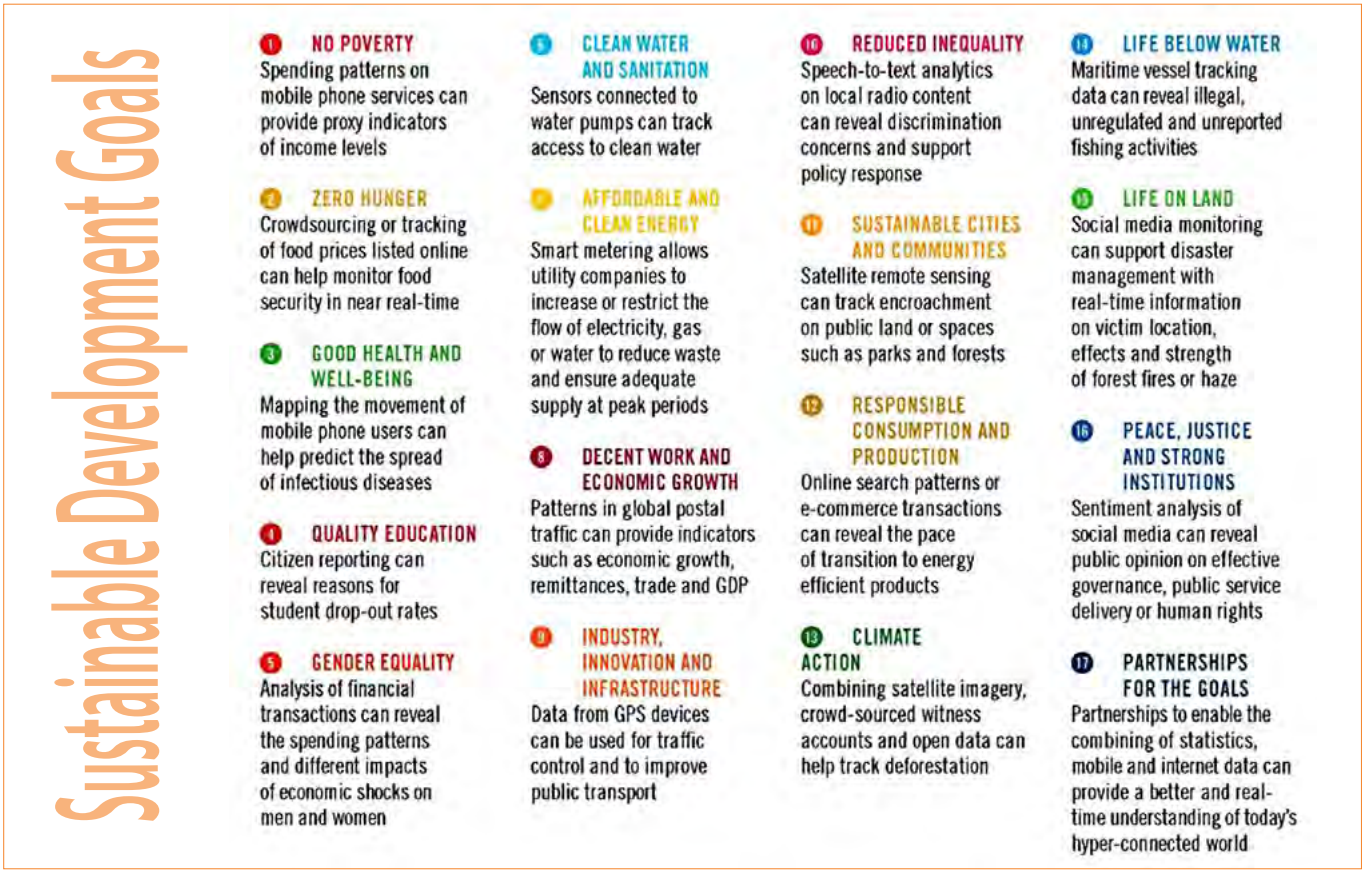
Big-Open data for people, society and the environment



Harnessing Big Data for Sustainable Development

Big Data from sources such as satellites, crowdsourcing, social media, sensors, GPS and mapping, aims to accelerate sustainable development and humanitarian actions.

Big Data applied to SDG



Source: http://unglobalpulse.org/sites/default/files/big%20data%20and%20sdgs.jpg

United Nations Data Ecosystem

As the data world is changing at a bewildering pace the United Nations (UN) works to

strengthen data ecosystems, develop global data principles, and integrate new technology with new partnerships. The UN Data Ecosystems aims to:

- link people, communities, businesses and governments
- integrate local, national and global data systems
- provide better data from a wider range of reliable sources e.g. multilateral institutions, civil society organisations, research institutions, academia, private sector and citizens
- design a framework for collecting, processing and analysing data
- use quality data from geospatial technology and satellite imagery (Earth Observations)
- deliver an accurate snapshot of progress towards the SDGs goals and targets
- contribute to an improved understanding of global, national and local social, economic and environmental problems, and how they can be sustainably managed

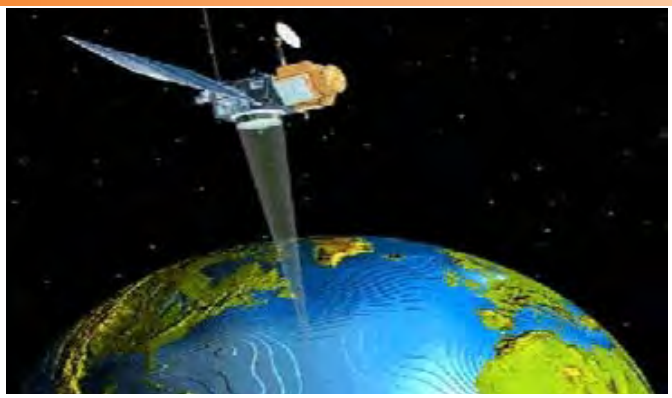
Taking the Global Pulse

Global Pulse is an initiative of the UN that attempts to bring real time monitoring and prediction to development projects. Mobile phones provide a powerful platform to reach remote people and obtain real-time feedback from communities. Global Pulse incorporates a diversity of data sources such as:

- interactive data visualisation tools used during a typhoid outbreak in Uganda
- extraction and analyses of tweets relating to vaccines and immunisation used in Indonesia
- exploration of online news data used for conflict analysis

Global Pulse has developed a partnership with Quid. Quid is an information mapping platform that enables visual mapping and analysis on most development topics.

PEOPLE & ECONOMIC ACTIVITY– PART 2: BIG DATA



Using Earth Observations-Big Data

Over 1,300 satellites are pivotal to our networked lives as they connect the world to broadband services, and its global positioning data (GPS), helps prevent and mitigate natural and human disasters.

'The Group on Earth Observations (GEO) is championing the cause of Open Data. "Whether it is from space, from the atmosphere, from the marine environment or from the land, if public resources and taxpayers' money went into building the satellites and/or the instrumentation, then the data needs to be released broadly and openly. Governments should not be charging their citizens for data that they have already paid for.'

Source: https://www.geospatialworld.net/article/geospatial-data-sustainable-development-goals/

'Earth observation data and information, collected from both space and ground platforms are the basis for addressing a multitude of environmental issues. A globally co-ordinated strategy for the collection, integration and sharing of these data and information plays an essential role in the post-2015 Sustainable Development Goals (SDG)'

Source: https://room.eu.com/article/Understanding_Earth_GEO_and_the_SDG_environmental_agenda

Earth observations provides critical data that assists SDGs, for example:

- SDG-2 (Zero Hunger) via crop monitoring
- SDG-6 (Clean Water & Sanitation) via population density, wastewater leakage data
- SDG-11 (Sustainable Cities & Communities) via air quality data
- SDG 14 (Life Below Water) via remote sensing for water quality
- SDG-15 (Life on Land) via forest cover data from satellites

SDGs and Earth Observations

Diagram: http://gefieo.org/sites/default/files/ieo/ieo-documents/SDG-Bigdata.pdf

Source: https://commons.wikimedia.org/wiki/File:SunBurst10.PNG

Geospatial Data monitors SDG

'Geospatial data, is the basis for evidence-based decision-making, monitoring and accountability. The geospatial community recognises that location and geography are significantly linked to many, if not all, elements of SDGs'

Source: https://www.geospatialworld.net/article/geospatial-data-sustainable-development-goals/

Geospatial Data links to SDGs

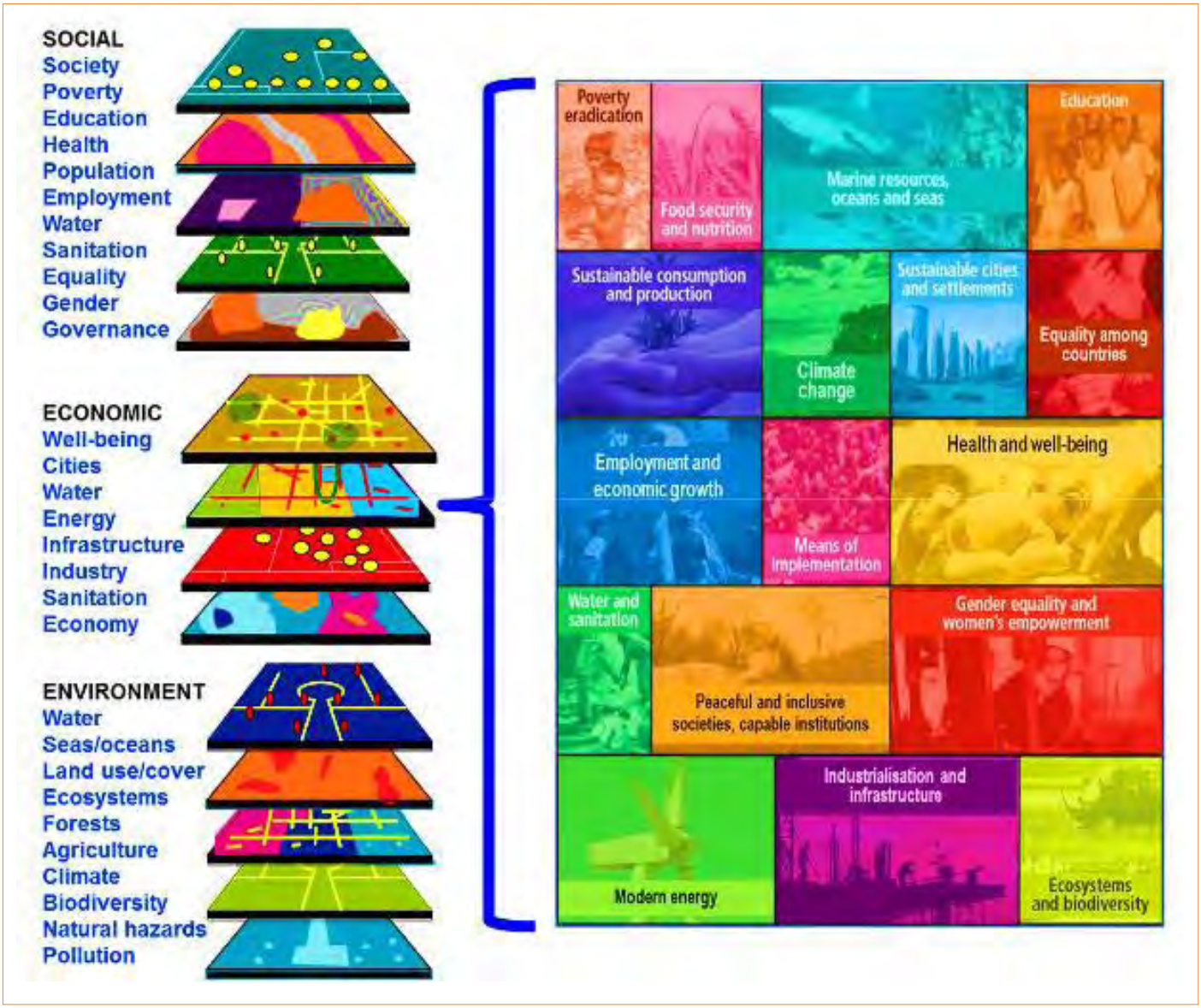


Diagram: https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-02/Statements/IAEG-SDGs-GGIM.pdf

Geospatial data aims to increase the availability of high-quality, timely and reliable data, disaggregated by geographic location to demonstrate differences in social, environmental and economic conditions, around the world. This data is critical to answering questions such as:

- Where is Ebola occurring? How do we contain the disease?
- Where are people at risk from rising sea levels? How do we protect these people?
- How many hectares of forests have been cleared in Indonesia for palm oil? Are these forest managed sustainably (economic, environmental, social)?

PEOPLE & ECONOMIC ACTIVITY– PART 2: BIG DATA

SDG Goals and use of Geospatial Data

SDG GOALS	11 CITIES	12 SUSTAINABLE CONSUMPTION	13 COMBATING CLIMATE CHANGE	14 MARINE AND COASTAL ECOSYSTEMS	15 TERRESTRIAL ECOSYSTEMS	16 PEACEFUL INCLUSIVE SOCIETIES
MAPS, GRAPHS AND DATA	• Public green spaces • Housing density • Substandard housing • Traffic congestion	• Energy use • Pollution – air, water • Landfills	• CO2 and methan emissions • Changes in sea levels, weather patterns (more frequent droughts, cyclones)	• Algal blooms • Eutrophication • Marine protected areas • Coastal erosion (natural and • manmade) • Clearing wetlands	• Land degradation • Salinisation • Biodiversity • Protected areas • Desertification • National parks • Wilderness areas	• Political violence • Fragile states • Crime hotspots • Refugee movements • Gross Domestic Product

Fighting climate change using Big Data and predictive analytics

The abundance of climate data from model simulations, satellites, geospatial data, and field observations is closing the knowledge gap on how the complex dynamic Earth system functions. As a result, climate science is one of the most data-rich domains in terms of data volume, velocity and variety. The cloud, the growing number of IoT and the rise of social media, resulted in additional data, photos and videos on anthropogenic changes to Earth. Continuous improvements to Big Data aims to lead to more accurate modelling of the global climate. Additionally, Data for Climate Action harnesses Big-Open Data from science and private and government sectors, and predictive analytics addresses the interconnected causes, impacts and mitigation of climate change.

Projects and predictive modelling

Many climate change projects are built around the principle of predictive modelling, with the most advanced climate models called General Circulation Models. However, sceptics question: How reliable are predictive climate change models? Is climate too complex to model or predict?

Innovative projects such as the International Centre for Tropical Agriculture (CIAT) project demonstrates how Big Data aids the mitigation of climate risks and strengthens resilience to climate disasters. In the past climate data provided weather forecasts, but when combined with crop data is able to deliver real time advice to farmers during droughts. Additionally the Weathersafe project designed to help humans deal with climate change, provides valuable data that assists coffee growers adapt to changing weather patterns and soil conditions.

Could the five Big Data projects stop climate change?

GOOGLE EARTH ENGINE	MICROSOFT'S MADINGLEY MODEL	DATA.GOV'S CLIMATE	GLOBAL FOREST WATCH	OPOWER
Tracks deforestation in the Amazon	Illustrates environmental impacts on animal mortality	Provides 400 government data sets from agencies and researchers	Tracks forest changes. Over 500,000 people use the service e.g. Nestle and Indonesian government	Analyses people's power usage

Race for technical superiority

Wining the digital race to achieve technical superiority over competitors involves gathering data to make sure you are not the next taxi company or hotel chain caught off guard by Uber or Airbnb.

Venture capitalist Marc Andreessen declared in:

- 2011 that 'software is eating the world' such as Uber is eating taxis and AirBnB eating hotels. 'Thanks to software, we no longer have to bring maps with us, we don't have to wait for the newspaper to be delivered at our doorstep, as all of our previously physical and bulky tools are now in our phones.'https://futurism.com/software-is-eating-the-world-and-you-need-to-see-it/
- 2017 that 'software is programming the world.' He predicted that in the near future 'chips will be free and be embedded in everything.'

Coding crystal ball

In the future detailed information from databases will be overwhelming! Apps such as Waze will track the flow of users, such as road workers blocking a lane. Autonomous cars will know the position of lampposts and cities will resurface roads when required. Streetlights will be replaced when they go out, and police will have more data on people walking along streets.

RealSense 3D Scanning

While the Intel RealSense 3D depth-sensing camera technology has applications for laptops, PCs and drones, it can also assist everyday tasks. By determining depth and scanning objects in 3D, the camera can verify whether the Ikea table will fit into a car's trunk. Travelers could use the RealSense camera's depth perception technology to scan carry-on luggage. This takes the guesswork out of determining if a bag meets the airline size requirements.

Source: http://www.cadalyst.com/%5Blevel-1-with-primary-path%5D/5-technologies-made-summer-splash-germany-25870

Talking Scooters

By saying 'hello, smart bike' into the helmet, people could sit on a BMW motorbike and inquire about the vehicle's route and tyre pressure. Using Intel Edison technology, the rider is able to access data from the bike's key computer system. Questions are interpreted by Intel voice-recognition software, and the answers are piped into the helmet's built-in audio system. This means that riders don't need to fiddle with a separate device or take their eyes off the road.

The back of the Intel Connected Helmet incorporates LED lights, showing all light signals from the scooter. The hope is that bikers will be more visible to drivers, which will reduce the number of road accidents

Source: http://www.cadalyst.com/%5Blevel-1-with-primary-path%5D/5-technologies-made-summer-splash-germany-25870

Big Data on wheels

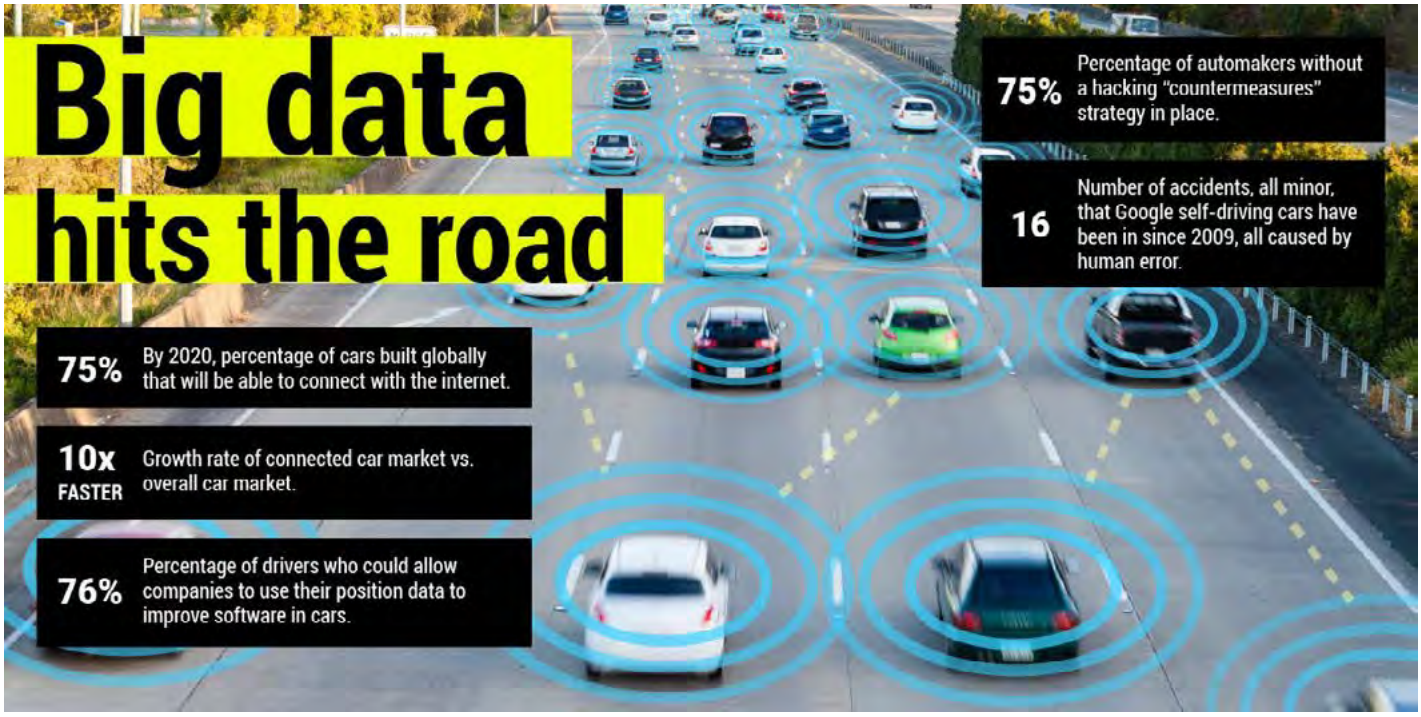
Modern cars are equipped with more than 100 sensors that create a constant stream of data.

Several times per second, sensors on cars measure location, performance and driving behaviour. According to a McKinsey report connected cars create up to 25 gigabytes of data per hour that is the equivalent of more than a month of 24-hour music streaming. However, cars equipped with surround cameras and radar are able to generate raw data of 100 gigabytes per second.

The data automatically generated by cars provides a huge source of Big Data and analytics. By 2018 one in five cars on the road will be self-aware and the global connected market will be worth $39 billion.

https://theconversation.com/why-driverless-cars-still-need-driving-tests-62721

Big Data hits the road



Statistics: http://static2.businessinsider.com/image/560052d069bedd9b0cfb6a2b-1200-600/connectedcars-1.png

Drive assist technology

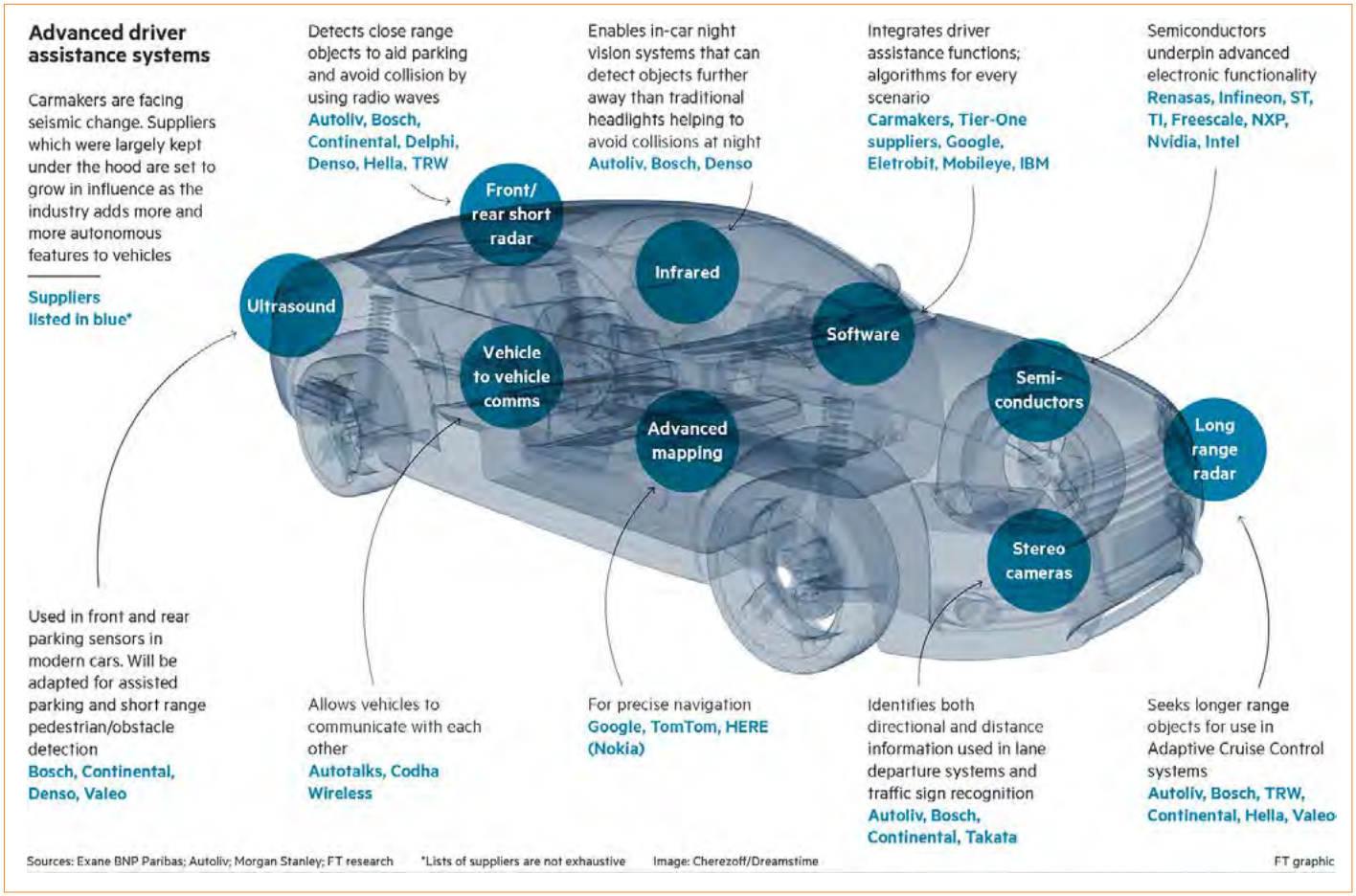


Diagram: http://im.ft-static.com/content/images/196878ca-5249-11e4-b55e-00144feab7de.img?width=2711&height=1770&title=&desc=

Merging Big Data with Artificial Intelligence (AI)

Data analysts are unable to keep pace with the huge quantities of Big Data created daily. This dilemma requires the use of Artificial Intelligence (AI), known as machine learning. AI adds an intelligence layer to Big Data, by assisting complex analytical tasks to be completed faster than humans.

AI is already revolutionising lives-Apple's Siri, Google's OK Google, and Amazon's Echo. For example, Siri is a computer program that operates as a personal assistant and knowledge navigator. It uses human languages to answer questions and perform actions by delegating requests to Web services such as looking for a restaurant or providing driving directions. AI also intends to deliver quicker drug discoveries, safer self-driving cars and alternative energy sources.

AI in Healthcare

AI applications will depend on data to develop predictive models. For example the larger the electronic medical data records reflect dangerous infections in hospitals, the better the system can predict these events before they occur. Referred to as actionable insights, AI used in healthcare aims to provide physicians with information to make better decisions for all patients

Healthcare using AI:

- Data management – collecting, storing, tracing lineage
- Design treatment plans for oncologists
- Assist radiologists detect health problems faster and more reliably
- Online consultations-report symptoms with checks against database and patient's history
- Health assistance and medication management
- Drug creation – speed process and make more cost effective

Image source: http://www.businesscloudnews.com/files/2016/01/AI-Artificial-Intelligence-Machine-Learning-Cognitive-Computing.jpg

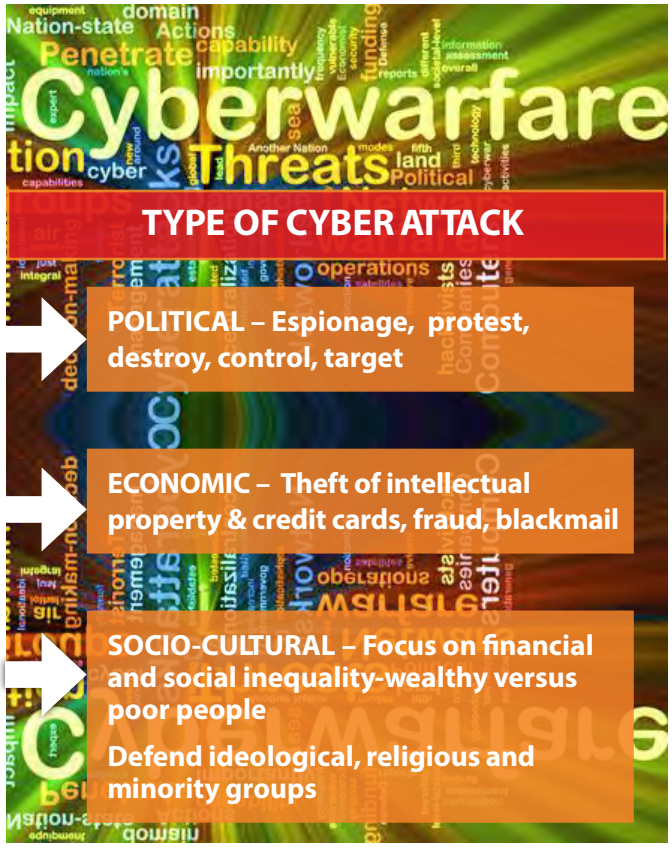
Internet of Things (IoT), Big Data and AI is crucial for smarter future decisions. A plethora of on-line free data platforms makes organising, synthesising and analysing data easier, however information in a database must be checked for errors, bias and duplications

Security threats to Internet of Things (IoT)

The Internet of Things (IoT), is an open system with many components, making it vulnerable and susceptible to security threats at the component level (e.g. mobile phone, PC) and the systems level (e.g. electricity supplies, banking).

As the number of connected devices is escalating, individuals, organisations and governments connected to the internet are increasingly susceptible to cyberattacks. This has led to the escalation of viruses, malware attacks, spear phishing and ransomware attacks. In May 2017, ransomware cyberattack or WannaCry, hit 200,000 victims in 150 countries. The ransomware locked computers and demanded payments between $A406 and $U812 to restore access.

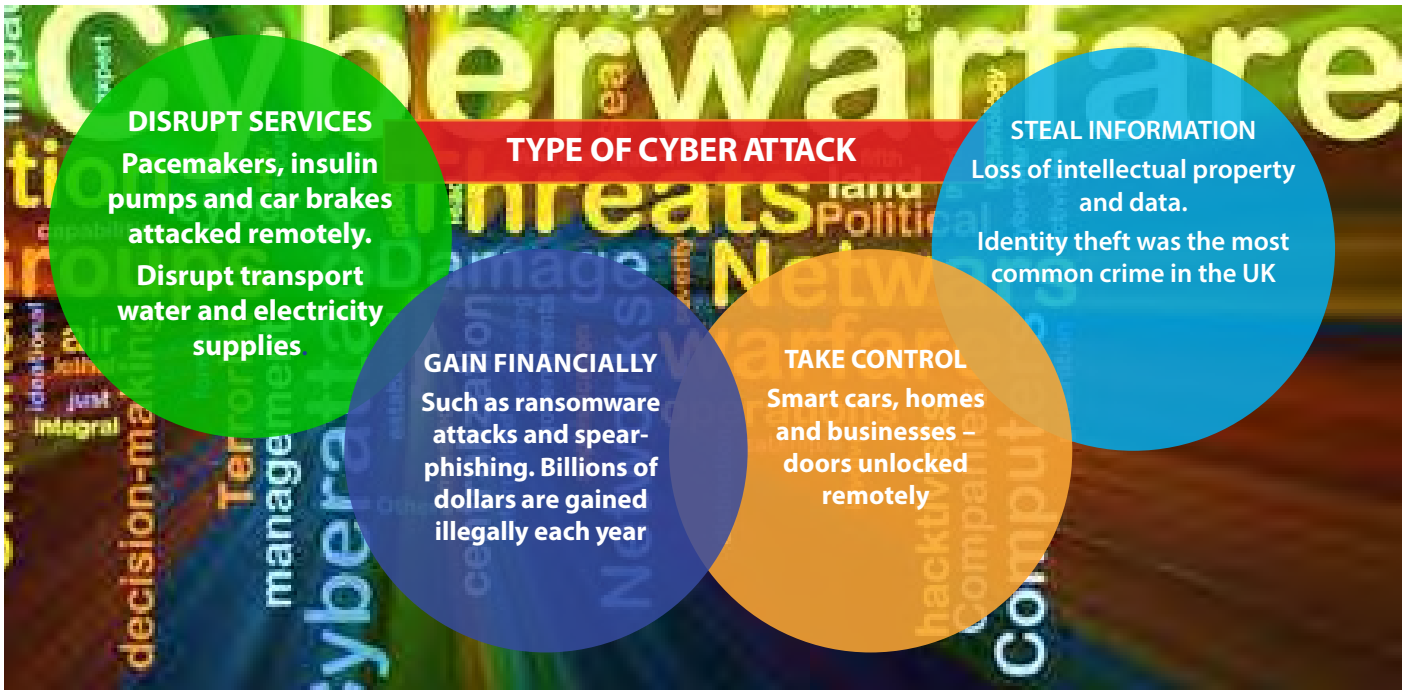
Types of cyber attacks



Background image: http://www.technewsworld.com/article_images/story_graphics_xlarge/xl-2015-cyberwarfare-1.jpg

Motivation of cyber attackers

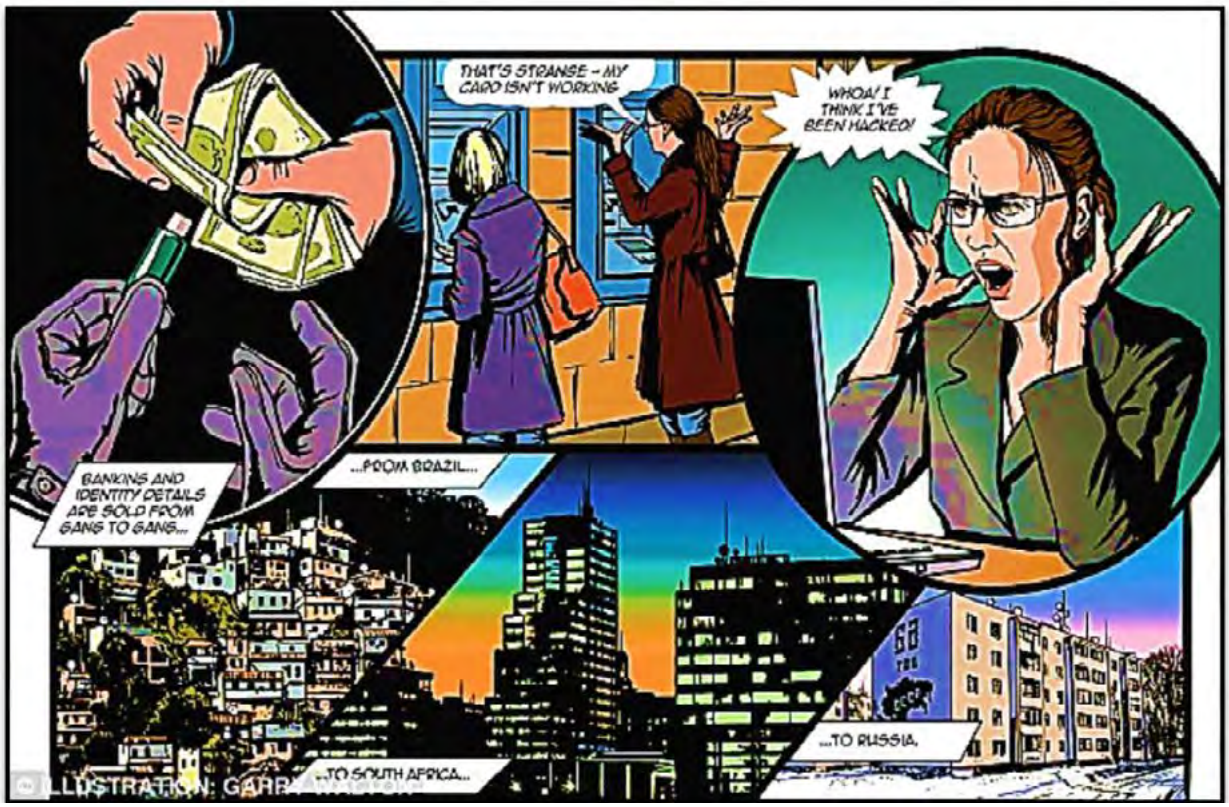
The cost of malicious cyber activity is mainly related to the theft of intellectual property and the loss of financial assets.



Hacker's tools

The hackers use a variety of tools such as sending personal e-mails, infiltrating baby monitors or smart TVs, eaves dropping on free public WiFi networks, impersonating trustworthy companies, sending bogus software updates, and using skimmers to steal ATM card information.

Stealing money via ATM



Cartoon: http://i.dailymail.co.uk/i/pix/2013/01/11/article-2260221-16E238A1000005DC-318_634x395.jpg

Security challenges

THREATS –
To privacy, mobile security and middleware security (computer software that 'glues together' separate and complex programs)

DANGERS – Of shared data or data stored in the cloud, runs risk of data and identity theft

ALTERNATIVE CHOICES

Convenience versus cost of security – Pre-programmed cars with auto unlock doors increases the possibility of manipulation for theft

Access to data versus security, privacy and transparency – Transparency reveals information organisations possess about a person

Image source: http://www.rodin.com.au/wp-content/uploads/2015/06/attack.jpg

Cybercrime

Cybercrime activities are globally diffused as it is open to everybody driven by profit and personal gain. Cybercrime has no boundaries although it is influenced by national laws and by efficiency of law enforcement. To counter the increase in cybercrime globally requires the adoption of defence mechanisms aimed to build a security culture.

Cybercrime activities affects IT. Cybercriminals leverage of Big Data has increased the effectiveness of attacks and disruption of businesses and governments. Web attacks and 'insiders' account for more than 55% of cybercrime costs per organisation. The highest cybercrime costs were in organisations such as defence, financial services, energy and utilities. Of greatest concern are the cybercrime acts represented by computer content, including child pornography, terrorism and piracy.

The number of crimes based on mobile devices and social media is exploding. In the majority of cases, the systems are exposed to cyber threats due to bad habits and risky behaviour. Bogus social network 'likes' and Instagram followers are sold to cybercriminals. The black market supports the growth of cyber threats within the cybercrime ecosystem

Cybercriminal ecosystem

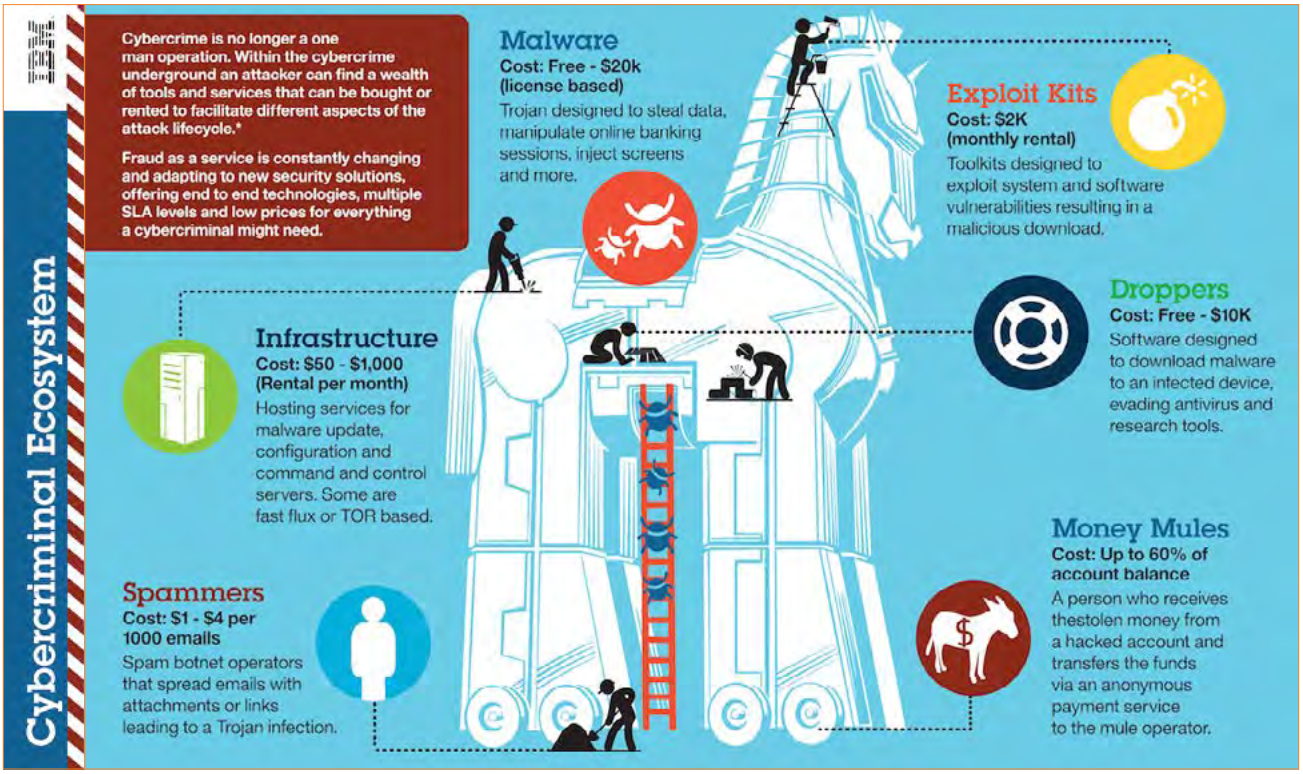


Diagram source: https://securityintelligence.com/wp-content/uploads/2015/06/Cybercrime-Ecosystem-Infographic-Final.jpg

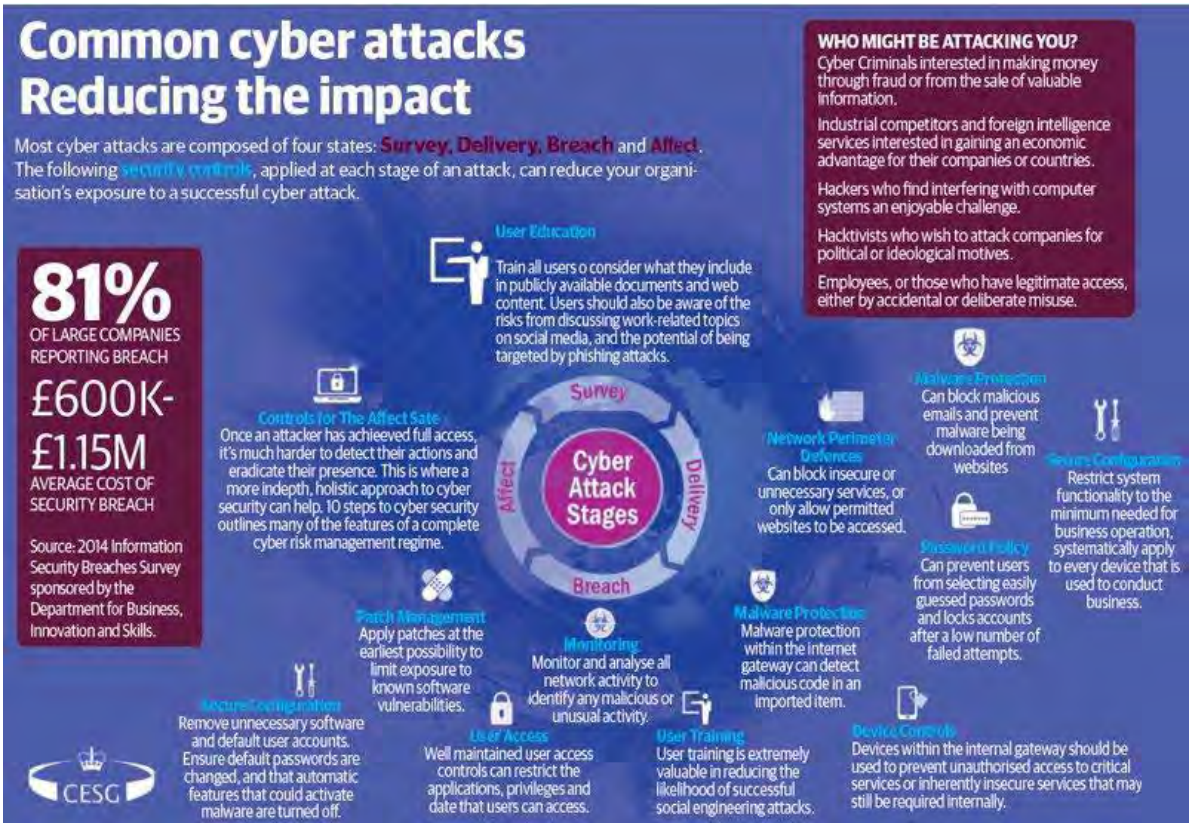
Reducing cyber attacks – the growing menace

Cyber attacks and hacking are threats to individuals, governments, businesses and international organisations. These attacks are generally for social, economic, political and defence purposes. In 2016, the largest attacks focused on financial services, governments and tourism, with cyberattacks on Dropbox, Linkedin and Yahoo.

Security controls and processes have been established to protect organisations and individuals from cyber attacks, such as:

- installing current antivirus software
- using malware protection to block malicious emails
- developing a password policy that prevents users from selecting easily guessed passwords
- monitoring and analysing abnormal network activity and data leakage
- ensuring data is regularly backed up and stored off-line

Common cyber attacks – reducing the impact



Source: http://www.dhakatribune.com/assets/uploads/2016/10/10-10.jpg

Cartoons on cyber attacks

Source: http://www.dezynetek.com/humour/cyberattack.jpg

Source: https://cdn1.lockerdome.com/uploads/edebf2c30150a767144ba2bcff8dc0b88d417683b7d84ecd0ffb3135a48b1314_large